

# Use of Twins to Study Environmental Effects

by Joe C. Christian\*

Extrapolation from pharmacogenetic studies would indicate that there is a great deal of genetic variability in the response of humans to noxious environmental agents. Co-twins provide the most closely matched genetic controls possible and, in addition, are matched perfectly for age and often have shared very similar environments since before birth. The efficiency of co-twin control studies can be further increased by the use of sequential analysis, so that for the studies of the effects of environmental agents on human populations, twins would give answers that would require many more unrelated subjects. Use of twins in epidemiology studies is now in its infancy, but investigators should carefully consider the use of this powerful experimental tool and begin to identify twins in population surveys.

## Introduction

Human experimentation is faced by ever-increasing costs, more competition for available resources and the ever-present need to protect the experimental subjects from unnecessary risks. To counter these forces, extreme care must be taken to select the most powerful and efficient experimental designs for human epidemiological research.

This paper explores the co-twin control model for use in epidemiological studies as an extremely efficient method of selecting controls which are matched genetically and environmentally as well as being the same age. Coupling paired comparisons of twins with sequential analysis is proposed as the most efficient method of studying low dose effects of environmental agents on human populations.

## Background

Twins have been used in human experimentation for over one hundred years. The comparison of monozygotic (MZ) twins with dizygotic (DZ) twins has been used frequently to estimate genetic variance or heritability. The MZ within-pair mean

square, for example, contains no genetic variance because MZ twins are genetically identical. The DZ within-pair contains one-half of population additive genetic variance. The difference between these mean squares ( $M_{WDZ} - M_{WMZ}$ ) is, therefore, an estimate of one-half population genetic variance. An area in which twin studies have been used almost exclusively to estimate genetic variance is the field of pharmacogenetics. Propping (1) reviewed studies of 11 drugs, all studied in twins and all showing a large amount of genetic variability in the rate of metabolism.

Brewer (2) coined the term, ecogenetics, by generalizing that genetic variation not only was relevant to drug action but needed to be considered in the human response to any kind of environmental agent. Because most of the environmental agents of concern have been implicated in teratogenesis, mutagenesis, carcinogenesis or the genesis of other diseases it is not feasible in many instances to do experimental studies of these substances in man. Therefore, observational or epidemiological studies are of prime importance.

Twins have just begun to be used in epidemiological studies. Leadership in this field has been taken by the National Heart, Lung and Blood Institute's Epidemiology Branch and the collaborative study of heart disease risk factors (3) by using the National Academy of Sciences National Research Council panel of World War II veteran twins (4).

---

\*Department of Medical Genetics, Indiana University School of Medicine, 1100 West Michigan Street, Indianapolis, Indiana 46223.

# Use of Twins in Epidemiological Studies

For most traits of importance in medicine, there is marked variability among individuals. In the most simple terms, this variability can be ascribed to genetic factors, environmental factors or an interaction between genes and environment. This inherent variability often obscures the effects of environmental influences by overwhelming them with background noise or experimental error and makes the choice of appropriate controls critical. Table 1 shows how co-twin control studies can be used to control much of this experimental error. The MZ within-pair mean square, for example, contains no genetic variance, which precludes the possibility of genetic-environmental variation and contains only a fraction of population environmental variance, inversely related to the environmental covariance of MZ twins. Because the twins share the same environment throughout gestation and often very similar environments after birth, this covariance often makes up a considerable portion of environmental variance making MZ twins often "as alike as two peas in a pod." Dizygotic twins, while not genetically identical, may also be used in co-twin control studies, and, while less concordant than MZ twins, are in most cases much more alike than age- and sex-matched controls often used in epidemiological studies. Dizygotic twins are also useful in studying the associations between genetic markers and disease states or risk factors (6).

There are several experimental situations for which co-twin control studies are suited. Historically, the most commonly used co-twin control study is a clinical trial in which the two members

of twin pairs are assigned randomly to two treatments or a treatment and a control. Twins are especially suited for long-term clinical trials where environmental trends are important and for studies in which carryover effects preclude the use of a crossover design. A second potential use for twins is in population surveys in which exposure history to an environmental factor is to be correlated with a disease endpoint. Twins can be ascertained as part of any population survey and then an appropriate sample chosen based upon exposure history with the co-twin serving as a built-in control.

In any proposed co-twin control study the added expense of obtaining twins must be compared to the expected advantage of using twins. It is often relatively inexpensive to ascertain twins which account for about 1% of the births in this country and therefore, one in fifty individuals. In addition, a recent review (7) revealed 21 active twin panels in the United States and Canada in addition to the resources for zygosity diagnosis available in virtually every state through clinical genetics units (8). The simple procedure of adding the question, "Are you a twin?" to population surveys would often provide a wealth of responses from which to select appropriate twin pairs for study.

The number of twin pairs needed to test a hypothesis can be calculated given an estimate of within twin-pair variability for the trait in question and the size of the difference to be detected. A method is available for estimating the number of twin pairs needed by using uniformity trials in which data from twin pairs not subjected to treatment are used to estimate the experimental errors of a completely randomized clinical trial and a co-twin control study (9). The within-pair variation ( $M_{WMZ}$ ) is an estimate of the co-twin control

Table 1. Analysis of variance model for twin studies.<sup>a</sup>

Source of variation	df	Mean squares	Expected value of mean square
Monozygotic twins			
Among pairs	$n_{MZ}-1$	$M_{AMZ}$	$2\sigma_a^2 + 2\sigma_d^2 + 2\sigma_i^2 + \sigma_e^2 + 4\sigma_{ge} + C_{MZ}$
Within pairs	$n_{MZ}$	$M_{WMZ}$	$\sigma_e^2 - C_{MZ}$
Dizygotic twins			
Among pairs	$n_{DZ}-1$	$M_{ADZ}$	$3/2\sigma_a^2 + 5/4\sigma_d^2 + (1+f)\sigma_i^2 + \sigma_e^2 + 2(\sigma_{ge} + \sigma_{ge}^*) + C_{DZ}$
Within pairs	$n_{DZ}$	$M_{WDZ}$	$1/2\sigma_a^2 + 3/4\sigma_d^2 + (1-f)\sigma_i^2 + \sigma_e^2 + 2(\sigma_{ge} - \sigma_{ge}^*) - C_{DZ}$

<sup>a</sup> $n_{MZ}$  = number of monozygotic twin pairs;  $n_{DZ}$  = number of dizygotic twin pairs; df = degrees of freedom;  $\sigma_a^2$  = variance component due to additive genetic effects;  $\sigma_d^2$  = variance component due to dominant genetic effects;  $\sigma_i^2$  = variance component due to epistatic genetic effects;  $\sigma_e^2$  = variance component due to environmental effects;  $\sigma_{ge}$  = covariance between genetic and environmental effects in the same individual;  $\sigma_{ge}^*$  = covariance between genetic effects on one member of a twin pair and environmental effects on the other member of that twin pair;  $C_{MZ}$  = covariance among environmental effects between pairs of monozygotic twins;  $C_{DZ}$  = covariance among environmental effects between pairs of dizygotic twins; and  $f$  = one minus the fraction of epistatic variance manifest within dizygotic twin sets. After Christian et al. (5).

## SUM OF TWIN-PAIR DIFFERENCES

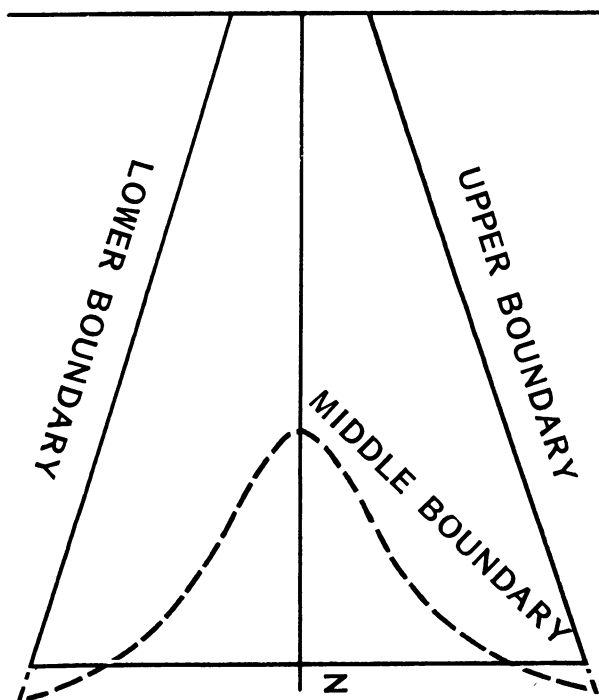


FIGURE 1. Schematic representation of a restricted sequential design for normally distributed measurements with known variability (12).

study experimental error. The among- and within-pair mean squares are used to obtain an estimate of experimental error for single-born individuals in the same population [ $(M_{AMZ} - M_{WMZ})/2$ ]. In this paper a nomogram is presented that allows the investigator to calculate the relative numbers of unrelated individuals and twins needed. This nomogram takes into consideration the relative sizes of the experimental errors and the relative degrees of freedom because a co-twin control study, like any paired comparison, has fewer degrees of freedom than a completely randomized design. A comparison of the relative efficiencies of matched pair studies with co-twin control studies is simpler because the degrees of freedom are comparable and the only additional information needed is an estimate of how much variability is controlled by the matched pairs as compared to twins plus an estimate of the difficulty of obtaining matched controls versus twins.

It has recently been proposed that another magnitude of efficiency can be obtained by coupling co-twin control studies with sequential analysis (10). Sequential analysis was developed in the early 1940's and its use in clinical trials has been

reviewed by Armitage (11). In sequential analysis the results are analyzed continuously and a decision made whether to continue or discontinue the trial with each analysis. This method has extreme appeal for human studies which are often conducted serially as appropriate subjects are ascertained. Figure 1 shows a model of a sequential analysis after Schneiderman and Armitage (12). This model is for normally distributed measurements with known variability and with no prior expectation of which treatment will result in the greater values for the variable being studied. The number of twin pairs to be studied ( $N$ ) is first estimated, and then twin pairs are studied serially with one member being assigned at random to each of two treatments. The sum of the differences is plotted (treatment A minus treatment B), and if this sum crosses the upper boundary, treatment A is judged to cause a higher value than treatment B. The reverse is true if the lower boundary is crossed. The upper and lower boundaries are generally set at  $\alpha$  levels of 0.05 or 0.01. Crossing the middle boundary, generally set at a much lower  $\alpha$  value than the upper or lower boundaries, indicates that continuation of the experiment gives little possibility of rejecting the null hypothesis. Therefore, crossing any of the boundaries results in termination of the experiment. It is not unrealistic to expect that on the average sequential analysis would reduce the number of experimental subjects needed by one half. In general, sequential analysis will add maximum efficiency in short-term studies or longer studies in which experimental subjects are ascertained over a relatively long period of time. There are series of sequential models well suited to various experimental conditions. If the study involves, for example, a population survey of environmental exposure and its relationship to another quantitative variable, the appropriate statistic could be the correlation between twin differences for the environmental influence and the twin differences for the trait being studied. In addition sequential designs are well suited for use qualitative traits that are either present or absent.

## REFERENCES

1. Propping, P. Pharmacogenetics. *Rev. Physiol. Biochem. Pharmacol.* 83: 123-173 (1978).
2. Brewer, G. J. Human ecology, an expanding role for the human geneticist. *Am. J. Hum. Genet.* 23: 92-94 (1971).
3. Feinleib, M., Garrison, R. J., Fabsitz, R., Christian, J. C., Huber, Z., Borhani, W. D., Kannel, W. B., Roseman, R., Schwartz, J. T., and Wagner, J. O. The NHLBI twin study

- of cardiovascular disease risk factors: methodology and summary of results. *Am. J. Epidemiol.* 106: 284-295 (1977).
4. Jablon, S., Neel, J. V., Gershowitz, H., and Atkinson, G. F. The NAS-HRC twin panel: Methods of construction of the panel, zygosity diagnosis and proposed use. *Am. J. Hum. Genet.* 19: 133-161 (1967).
  5. Christian, J. C., Kong, K. W., and Norton, J. A. Choice of an estimate of genetic variance from twin data. *Am. J. Hum. Genet.* 26: 154-161 (1974).
  6. King, R. J., Feinleib, M., Garrison, R. J., and Christian, J. C. Studies of associations between genetic markers and plasma lipids in twins. In: *Progress in Clinical and Biological Research*, Vol. 24c, Alan R. Liss, New York, 1978.
  7. Christian, J. C. The use of twin registers in the study of birth defects. *Birth Defects: Orig. Art. Ser.* 14(6A): 167-178 (1978).
  8. Lynch, H. T., Fain, P., and Marrero, K. *International Directory of Birth Defects Genetic Series*, 6th ed. March of Dimes Birth Defects Foundation, White Plains, N. Y., 1980.
  9. Christian, J. C., and Kang, K. W. Efficiency of human monozygotic twins in studies of blood lipids. *Metabolism* 21: 691-699 (1972).
  10. Christian, J. C., Yu, P. L., Grim, C. E., and Norton, J. A., Jr. The use of sequential analysis to improve the efficiency of co-twin control studies. *Acta Genet. Med. Gemellol.* in press.
  11. Armitage, P. *Sequential Medical Trials*. John Wiley & Sons, New York, 1975.
  12. Schneiderman, M. A.; and Armitage, P. A family of closed sequential *t*-tests. *Biometrika* 49: 359-366 (1962).